An Analysis of Diffusion of Teacher-curated Resources on Pinterest

Hamid Karimi
Utah State University
hamid.karimi@usu.edu

Kaitlin Torphy Knake
Michigan State University
torphyka@msu.edu

Kenneth A. Frank
Michigan State University
kenfrank@msu.edu

ABSTRACT

Teachers increasingly rely on online social media platforms to supplement their educational resources, greatly influencing PK-12 education through the swift and extensive diffusion of teacher-curated resources. Understanding this diffusion process is crucial, but current educational studies primarily report resource diffusion through small-scale analyses, such as teacher interviews or anecdotal accounts. To bridge this gap, we conduct a pioneering, large-scale, quantitative, and data-driven analysis of the diffusion of teacher-curated resources on Pinterest, a platform widely embraced by educators. Our study begins by defining a resource's diffusion tree, which encapsulates the cascade of resource sharing across the social network. Based on this diffusion tree, we identify three measures to characterize a resource's diffusion process: volume, virality, and velocity. Equipped with these three measures, we conduct an in-depth analysis of the diffusion of over one million resources curated by thousands of teachers on Pinterest. Our investigation concludes by examining the correlation between a resource's attributes and its curator's attributes and the diffusion of the resource.

Keywords

Teachers, Social Media, Diffusion, Pinterest, Education

1. INTRODUCTION

Historically, teachers expanded their knowledge base through formal and informal professional development channels. They formed networks through direct, face-to-face interactions with peers and, more recently, through online communities for exchanging knowledge, experiences, and social capital. The emergence of online social media platforms like Facebook, Twitter, and Pinterest has provided teachers with a new platform for connecting with like-minded peers and sharing pedagogical resources. This phenomenon has stimulated a surge in academic studies focused on teachers' engagement with social media [1, 2, 3, 4, 5, 6]. Contrasting traditional methods of educational resource curation, which can be time-consuming and scale-limited, the accessibility of sourcing educational resources from fellow teachers on social media platforms has become highly appealing. Teachers can now readily access materials from those they admire or perceive as field experts.

Additionally, the diffusion of these resources can occur swiftly, often within the same day, enabling teachers to integrate new materials into their classroom practices efficiently. Across social media, the established social networks and professional communities of teachers have facilitated the diffusion of information and instructional resources on an unprecedented scale [7]. Previously, teachers might have had only a handful of colleagues to turn to for advice or information. Now, they can access a broad spectrum of instructional resources and interact with "teacherpreneurs" from across the globe [8]. Consequently, the fast and efficient diffusion of resources has become a new norm, significantly influencing pedagogical practices and educational dynamics.

While previous research has explored the diffusion of information among teachers, often referred to as the exchange of knowledge or resources [9], there remains to be a significant gap in our understanding of the large-scale propagation of teacher-curated resources on social media platforms. Specifically, investigations need to be more into how these resources navigate through the network and the influence of the resource attributes and its curator on this propagation process.

To address this, we conduct a comprehensive, large-scale analysis of the diffusion of teacher-curated resources on Pinterest, a platform popular among teachers [10]. We start by gathering a substantial sample of Pinterest-using teachers and detailed information about their curated resources. Subsequently, we construct the diffusion process for over one million teacher-curated resources on Pinterest. This process encapsulates several vital elements: the initial curator of a resource, subsequent users who have re-shared the resource, and the timeline of the resource's re-sharing.

These vital details about a resource's diffusion process are captured in a diffusion tree, as demonstrated in Figure 1, where we also display the pin curation time beneath each node (more about diffusion trees in Section 3.3). This example illustrates the speed of resource diffusion via social media, highlighting the platform's power in swiftly dissemi-

H. Karimi, K. T. Knake, and K. A. Frank. An analysis of diffusion of teacher-curated resources on pinterest. In M. Feng, T. Käser, and P. Talukdar, editors, *Proceedings of the 16th International Conference on Educational Data Mining*, pages 55–66, Bengaluru, India, July 2023. International Educational Data Mining Society.

© 2023 Copyright is held by the author(s). This work is distributed under the Creative Commons Attribution NonCommercial NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.

https://doi.org/10.5281/zenodo.8115713

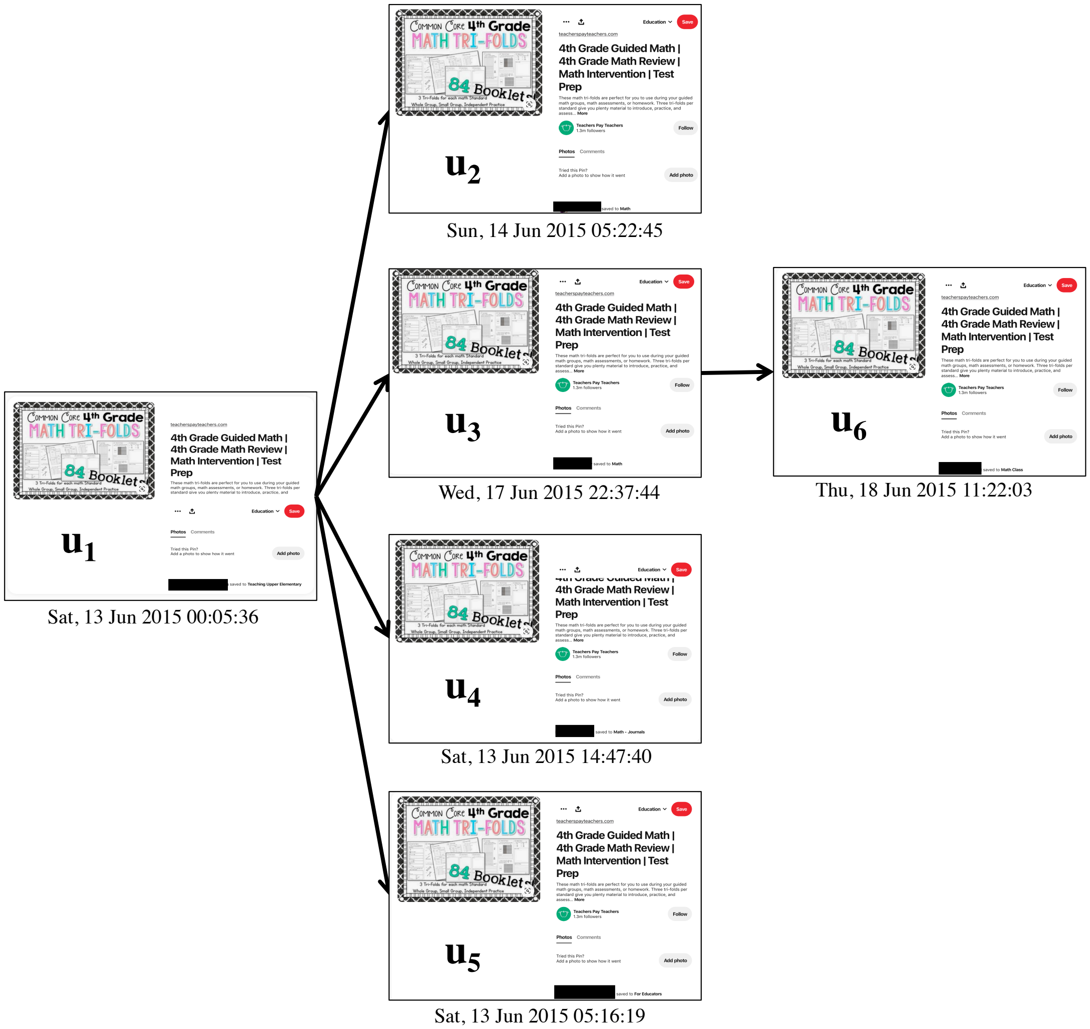


Figure 1: An example of a diffusion tree illustrating the prorogation of a teacher-curated resource on Pinterest

nating educational resources.

We then introduce three key measures that characterize the diffusion process: the number of users who have received a teacher-curated resource (volume), the structure-related penetration of a resource in the network (virality), and the speed of resource diffusion (velocity). Leveraging these measures, we conduct a large-scale analysis of teacher-curated resource diffusion and address two pivotal research questions. First, do resource attributes, such as their topics or sources, impact diffusion? Second, how do teacher-related attributes, such as the number of online followers, influence the diffusion of their curated resources?

This study's novel analysis and findings significantly contribute to the knowledge surrounding teaching and teacher learning with social media. Specifically, it helps illuminate how social media assists teachers in acquiring resources for their pedagogical practices. In summary, our contributions in this study are as follows:

- ➤ We construct the diffusion trees for over one million resources shared by thousands of teachers on Pinterest, offering a comprehensive visualization of resource propagation.
- ➤ We introduce three key measures - volume, virality, and velocity - to effectively characterize the diffusion of resources based on the constructed diffusion trees.
- ➤ We conduct a large-scale analysis of the diffusion of teacher-curated resources, outlining the relationship between the attributes of a resource and a teacher and how these relate to resource diffusion.

The rest of this paper is organized as follows. First, in Section 2, we present a brief literature review. Next, in Section 3, we discuss the dataset. Then, in Section 4, we introduce measures characterizing the diffusion process. Section 5 includes our analysis of the diffusion of teacher-curated resources. Finally, we conclude the paper in Section 6.

2. RELATED WORK

Online social media platforms offer significant benefits to teachers, notably in the domain of instructional resource curation [3]. Pinterest, an image-based personalized social media platform, is pivotal in this regard, boasting 440 million active users per month [11]. American teachers widely adopt it as a professional platform and a virtual repository of resources [8, 12, 13]. A national survey by the RAND Corporation underscores this trend, revealing that most elementary and secondary teachers in the United States utilize Pinterest to cater to their instructional needs [10].

The qualitative analysis conducted by the authors in [13], based on interviews with eight teachers, sheds light on the functionality of Pinterest in the educational sphere. They recruited teachers through snowball sampling on Twitter and found that educators viewed Pinterest as a digital organizer, compiling resources they discovered online or developed themselves. This echoes findings from previous studies that emphasized Pinterest's role as a content curation tool [14, 15, 16, 17].

Further exploring this theme, Schroeder et al. [18] conducted a qualitative study involving 117 teachers and found that educators predominantly used Pinterest to find resources tailored to their classroom requirements. Moreover, Torphy Knake et al. [8] investigated teacherpreneurial behaviors on Pinterest. After analyzing the source of 140,287 resources curated by 197 teachers, they found that educational blogs were the primary origin of these resources. In addition, market websites specifically targeting teachers, notably teacherspayteachers.com, also contributed significantly to the source of pins.

Additionally, their study revealed that a substantial majority of pins (82.8%) were monetized. Hu et al. [12] examined the curation mechanism of mathematical resources on Pinterest, discovering that these resources typically exhibited low cognitive demand. Their research also demonstrated the role of socialized knowledge communities in assisting mathematics teachers in finding relevant resources. Lastly, [19] provided insightful analysis into the curation practices of mathematical resources, identifying three types: self-directed, incidental, and socialized. A key takeaway from their study is their insight into how Pinterest-sourced educational resources are utilized in the classroom.

The work most closely related to our study is that by Liu et al. [20], which examined the process of Pinterest resource curation among 34 early career teachers (ECTs) from three states in the Midwest. They focused on the diffusion of resources among an ECT and their colleagues within the same school, whom the ECT nominated as close colleagues. Their findings suggest that Pinterest serves as a conduit between weakly connected teachers in the same school.

However, our study presents several significant improvements compared to [20]. Firstly, we operate on a much larger scale, investigating the diffusion of over one million resources among thousands of teachers. Secondly, while their study examined diffusion through a single direct re-pinning between two teachers, we delve into the entire cascade of information diffusion as represented by the diffusion trees. Thirdly, their study was limited to teachers within the same school who have potential face-to-face interactions. In contrast, we examine diffusion among teachers on an online platform without consideration for potential real-life interactions.

3. DATASET

This section provides an overview of the dataset we utilized for our study. We detail the process of teacher sampling, explain our approach to automatic teacher identification, and illustrate how we construct the diffusion trees.

3.1 Teacher Sampling

As a part of an interdisciplinary project called "Teachers in Social Media"[1], we surveyed 540 teachers across five U.S. states: Illinois, Indiana, Michigan, Ohio, and Texas. We then harnessed the Pinterest API (Application Programming Interface) to gather data about these surveyed teachers and their online connections, including followers and followees. The collected data for each user encompasses their pins and boards. Every pin carries an image (or, in recent times, a video), a description, a title, a link to its source, a board, the parent pin, and other supplementary information. The parent pin refers to the preceding pin from which the current pin has been re-pinned (re-shared). A board is a user-generated catalog that organizes pins with similar themes (for instance, all pins related to 'multiplication table instruction').

3.2 Automatic Teacher Identification

As stated earlier, the principal aim of this paper is to conduct a large-scale analysis of the diffusion of resources curated by teachers. However, utilizing data from only the surveyed teachers would not suffice to accomplish this goal, as we have surveyed a relatively small number of teachers. Therefore, one might suggest increasing the number of surveyed teachers. However, surveying is a time-consuming and expensive process. As such, developing a method capable of identifying teachers *automatically* becomes highly beneficial, especially considering that we have already collected data from thousands of users connected to the surveyed teachers. Moreover, based on the principle of homophily (i.e., the tendency for individuals to associate with others similar to themselves [21]), which is prevalent in (online) social networks, it is highly probable that a significant portion of the surveyed teachers' online connections are indeed teachers.

Fortunately, in our prior study [5], we introduced a machine learning-based method capable of efficiently and effectively identifying teachers on Pinterest. For reference, Figure 2 provides a comprehensive view of our previously proposed method. The input for this method is the data of an unlabeled user (i.e., an online friend of a surveyed teacher), and the output is the probability that this user is a teacher, denoted as p. We establish a threshold τ; if $p > \tau$, the user is considered a teacher; otherwise, they are classified as a non-teacher. Employing a conservative threshold of $\tau = 0.9$, we automatically identified approximately 16,000 additional teachers. Our rigorous evaluation of this method in our previous study indicated a minimal error in teacher classification. Specifically, we conducted an exhaustive resiliency

[1] `https://www.teachersinsocialmedia.com/`

analysis of this method, ensuring it is a robust and reliable approach for automatic teacher identification on Pinterest.

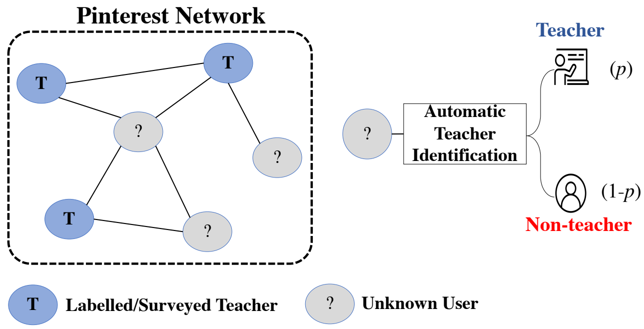


Figure 2: An overall illustration of our previously developed automatic teacher identification method

Table 1: Basic statistics of our constructed dataset

#Users (teachers)	**13,267**
#Pins	**1,162,983**
#Boards	**865,655**
#Followees	**11,84,940**
#Followers	**1,046,729**

Table 1 presents basic statistics of our compiled dataset. As the table indicates, our dataset comprises 13,267 teachers, who were either surveyed directly or identified automatically using our method. Furthermore, these teachers have curated over one million pins.

3.3 Diffusion Trees

We constructed diffusion trees for these resources to investigate the diffusion of curated resources on Pinterest. A diffusion tree is a directed graph, symbolized as $T = (U, E, p, r)$, representing the cascade of user information. Here, U denotes the set of users engaged in the diffusion, E is the set of directed edges between users in U, p is the pin being disseminated among users in U, and r is the origin or root of the tree—a teacher who initially curated the pin p. Each edge $e = (u_i, u_j) \in E$ suggests that the user $u_i \in U$ has received pin p from user $u_j \in U$ and subsequently re-pinned it. For instance, in Figure 1, user u_1 (the root) has curated a resource that has been disseminated throughout the network and re-pinned by users $u_2, u_3, u_4, u_5,$ and u_6.

We constructed diffusion trees for 1,162,983 unique pins that our identified teachers curated, meaning the root node of each tree was one of the teachers we identified, as described in Section 3.1. It is important to note that not all users in U were necessarily teachers. Furthermore, we created trees for all types of resources curated by teachers, both educational and non-educational. This was done for two main reasons. Firstly, including non-educational pins allows us to better contextualize the diffusion patterns of educational resources compared to non-educational ones. Secondly, apart from studying the diffusion of teacher-curated resources on Pinterest, a secondary objective of this paper is to analyze teachers' general behavior on the platform. Therefore, examining the diffusion of all types of teacher-curated resources contributes to this secondary objective. Lastly, it is worth mentioning that our dataset of diffusion trees represents the largest dataset of diffused resources on Pinterest to date, offering the potential for future research on information diffusion on social media.

4. DIFFUSION MEASURES

We present three measures to characterize the diffusion process, inspired by those introduced in [22]. These measures aim to evaluate the large-scale and fast diffusion of educational resources on social media, as documented in previous studies [12, 20, 23]. Specifically, these measures are designed to echo the two critical aspects emphasized in prior research on the diffusion of educational resources on social media, particularly Pinterest: a) educational resources are disseminated on a large scale among teachers, and b) this dissemination of educational resources occurs rapidly [20, 12].

4.1 Volume

The first measure, *volume* (VL), is defined as the total number of nodes in a diffusion tree:

$$VL(T) = |U| \tag{1}$$

For instance, the volume of the tree depicted in Figure 1 is 6. Despite its apparent simplicity, the volume measure carries significant implications as it indicates how much information has diffused. Specifically, the count of users that have received the information is used in predicting or assessing the popularity of information on social media [24, 25]. Relevant to our study, we can determine the level of interest other users or teachers have in a teacher-curated resource by examining its volume.

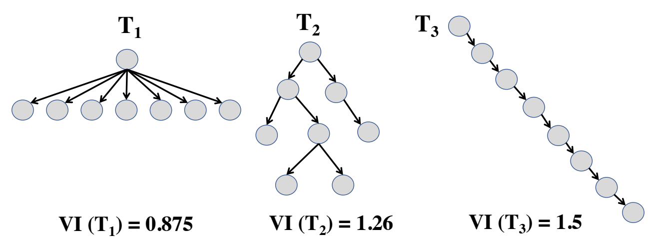
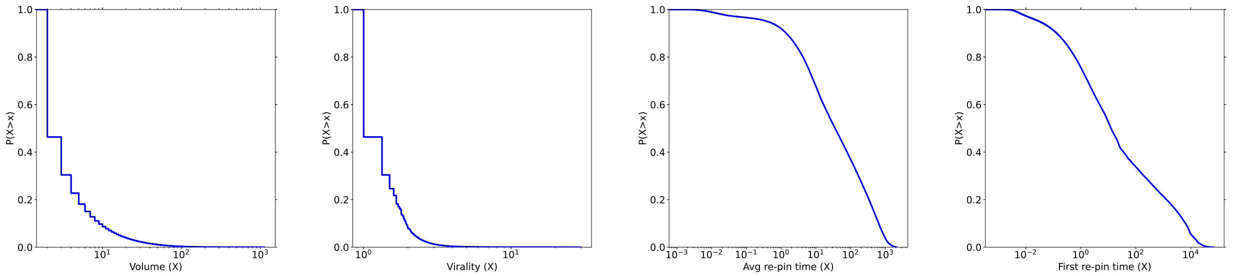


Figure 3: Three diffusion trees with the same volume but different virality values

4.2 Virality

While the volume measure is important, it only reports the number of individuals who have re-shared a resource. However, depending on the structure of a diffusion tree, the dissemination can take different forms. To illustrate this, Figure 3 presents three distinct diffusion trees, all having a volume of 8 but exhibiting very different forms of dissemination. In T_1, there is a broadcast from the root to other nodes, with only the root participating in the information propagation. In contrast, T_2 involves more nodes in the diffusion process. T_3 represents an extreme scenario with a chain-wise 'deep' tree, where the message has been passed on consecutively. Distinguishing between diffusion scenarios based on their tree structure provides insight into the virality and penetration of a message across the network [22].



(a) Volume (Eq. 1) (b) Virality (Eq. 2) (c) Average re-pin time (Eq. 3) (d) The first re-pin time (Eq. 4)

Figure 4: The CCDF plots of the defined diffusion measures (x-axes are in log scale)

Table 2: Some statistics of the introduced diffusion measures of the constructed diffusion trees

Diffusion Measure	Min	Max	Mean	Median	Std	top 0.1%	top 0.01%
Volume	2	1,129	5.4	2	13.58	> 174	> 434
Virality	1	29.72	1.33	1	0.54	> 5.99	> 11.45
ART	0.0012	2,159.4	192.4	35.8	317.3	$> 1{,}950.7$	$> 2{,}113.2$
FRT	0.0008	65,655.0	1,814.4	12.5	4,975.0	$> 45{,}020.7$	$> 56{,}960.9$

Therefore, we define the *virality* (VI) of a diffusion tree as follows:

$$VI(T) = \frac{2}{(|U|) \times (|U|-1)} \sum_{\forall u_i, u_j \in U} d(u_i, u_j) \qquad (2)$$

Here, $d(u_i, u_j)$ represents the shortest distance between two users u_i and u_j in the diffusion tree T. The sum of the shortest distances between nodes in a graph is known as the Wiener Index [26, 27]. The term $\frac{2}{(|U|)\times(|U|-1)}$ normalizes the Wiener Index. Based on this measure, we can observe that T_3 has the highest virality among the trees in Figure 3.

4.3 Velocity

Alongside volume and virality, the speed of diffusion is also crucial. Previous studies have highlighted the rapid diffusion of educational resources on social media, especially Pinterest, making these platforms highly appealing to teachers [28, 23]. Thus, our third diffusion measure pertains to the velocity (or speed) of diffusion. For this, we introduce two metrics.

The first metric is the *average re-pin time*, which calculates the average time between two re-pins in the diffusion tree. The average re-pin time (ART) for a diffusion tree is defined as:

$$ART(T) = \frac{1}{|U|-1} \sum_{\forall e \in T} u_j(t) - u_i(t) \qquad (3)$$

Here, (u_i, u_j) is an edge in the diffusion tree and $u_i(t)$ ($u_j(t)$) represents the re-pin time by user u_i (u_j). In Eq. 3, we have subtracted $u_i(t)$ from $u_j(t)$ as the user u_i received the pin earlier. Given the rapid diffusion of information on social media, we use an hour as the time scale. The ART for the example tree shown in Figure 1 is 46.2 hours.

However, sometimes a resource can continue to be diffused for an extended period (for example, months), which can result in a large ART. Therefore, to better capture the diffusion velocity, we define the *first re-pin time* (FRT). It represents the time duration from the initial curation of a pin to its first re-pin:

$$FRT(T) = min\{u_i(t) - r(t)\} \;\; s.t. \;\; (r, u_i) \in E \qquad (4)$$

Here, $r(t)$ denotes the time the root curated the pin. The FRT for the example tree in Figure 1 is 5.16 hours.

5. DIFFUSION ANALYSIS

In this section, we examine the diffusion trees we have constructed. First, in Section 5.1, we provide statistical data on diffusion measures. Next, in Section 5.2, we discuss how different resource types are diffused. Lastly, in Section 5.3, we explore the relationship between the earlier diffusion measures and specific attributes related to teachers.

5.1 Statistics of Diffusion Measures

This section delves into the statistics and distributions of the three diffusion measures. Table 2 provides specific statistics about virality, volume, and velocity measures. In addition, the CCDF (complementary cumulative distribution function) of the diffusion measures is depicted in Figure 4.

As depicted in Figures 4a and 4b, both volume and virality exhibit a power-law distribution. This suggests that while most resources have low volume and virality, a small percentage displays exceptionally high values for these measures. Further, as per Table 2, the top 0.1% of diffused resources exhibit a volume and virality exceeding 174 and 5.99 hours,

Figure 5: A sample of a popular pin from *moffattgirls.blogspot.com* account, a prolific educator, that has been received (re-pinned) by 936 other users

respectively. This indicates that certain resources curated by teachers have gained considerable popularity. These findings align with prior studies on the virality and popularity of information on social media, demonstrating that some information can become significantly viral across the network [29, 30, 22]. On average, approximately five users have re-pinned each teacher-curated resource on Pinterest.

Contrary to volume and virality, the velocity measures do not adhere to a power-law distribution as seen in Figures 4c and 4d. Furthermore, a notable disparity exists between the mean and median for ART and FRT. While the median average re-pin and first re-pin times are relatively short (35.56 and 12.56 hours, respectively), their means are skewed due to the presence of outliers.

In conclusion, teacher-curated resources diffuse rapidly and reach a significant number of other users on Pinterest, including other teachers. Although this observation has been suggested in anecdotal reports [20, 12], our study offers the first large-scale data-driven analysis to confirm it.

5.2 Resource Attributes and Diffusion

In this part, we explore the diffusion measures in relation to two key attributes of pins: their topics and domains.

5.2.1 Topic

Every pin on Pinterest is associated with a pre-defined topic (or category), such as *travel*, *education*, or *fashion*. Figure 6a presents the average volume value for each topic. As evidenced by this figure, pins categorized under *education* exhibit the highest volume, with each such pin being received by an average of six users on Pinterest. Interestingly, *kids* ranks second in terms of volume, a fact that could be partially attributed to this topic's similarity to *education* and its appeal to teachers, particularly for resources specific to pre-kindergarten or homeschooling. All other topics exhibit lower volumes, generally below 4. Given that the dominant topic is *education*, and there is limited data for other topics, these topics demonstrate relatively high standard deviations.

Figure 6b displays the average virality value for each topic. As with volume, *education* also records the highest virality, indicating the extensive penetration of teacher-curated educational resources across Pinterest. The topic *kids* also showcases a relatively high average virality value. Moreover, comparing the volume and virality values in Figure 6a and Figure 6b suggests that high volume does not necessarily equate to high virality. For example, pins categorized under *quotes* exhibit relatively high virality, but their volume is not as impressive.

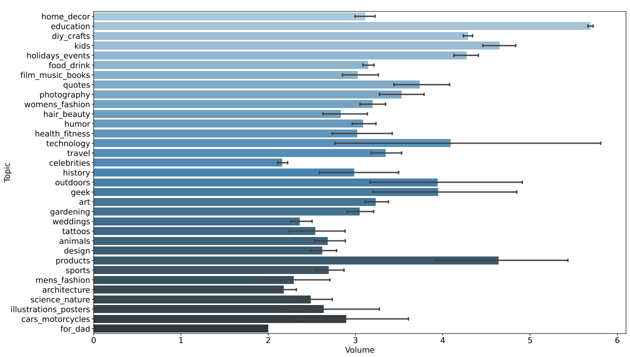
Figure 7 depicts the median values for the average and first re-pin times. We opted to use the median for these plots since, as discussed in Section 5.1, the ART and FRT values of our constructed diffusion trees include some outliers. Moreover, specific topics have limited data, resulting in skewed velocity measures. Therefore, for clarity, we present the median velocity measures for topics with a pin proportion of at least 10%.

From the results of the velocity measures, we observe two key points. Firstly, the topic *education* has both a short average re-pin time and first re-pin time. Specifically, the median of the first re-pin time for *education* is just 12 hours, suggesting that a teacher-curated resource takes roughly half a day to be received by another user on Pinterest. This underscores the rapid diffusion of educational materials across Pinterest. Secondly, the average re-pin time is generally longer than the first re-pin time. We posit that this may be due to a user quickly saving a pin curated by the root, with the pin then spreading across the network at a slower rate. However, there are a few exceptions to this, such as *travel* and *art*. This could be attributed to the unique appeal of these topics to teachers, whose pins may take some time to attract attention initially. However, once they gain traction, they diffuse more rapidly.

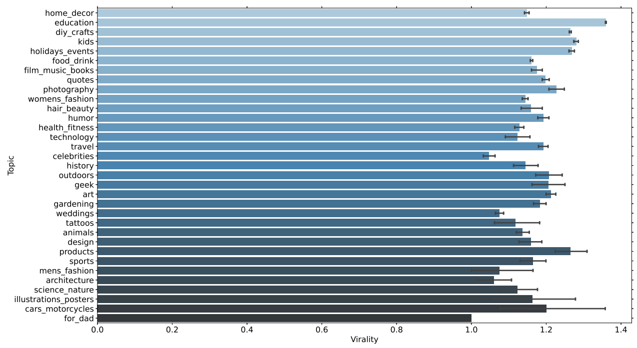
5.2.2 Domain

Pinterest allows users to pin resources from anywhere on the web. Given this property, examining the diffusion of pins from various sources becomes essential. In this section, we analyze the diffusion of teacher-curated resources based on the domains of their sources. For this analysis, we consider only the top 10 domains preferred by teachers. Figure 8 presents the average volume and virality values for these top 10 domains. Figure 9 displays the median of the average re-pin time and the first re-pin time for the exact top 10 domains. From these results, we draw the following observations:

- Except for *youtube.com* and *Uploaded by User* (resources directly uploaded from the user's device), the remaining domains are predominantly education-related, for example, *moffattgirls.blogspot.com*. Interestingly, pins from *moffattgirls.blogspot.com* record the highest volume. This blog is managed by a former elementary school teacher who exclusively creates educational materials. Further investigation reveals that this teacher is highly active and influential on *teacherspayteachers.com*– the largest online marketplace for instructional resources. Therefore, it is not surprising that her educational materials garner significant interest.

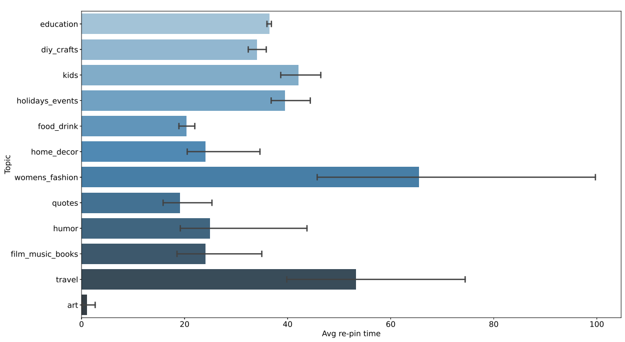


(a) Volume

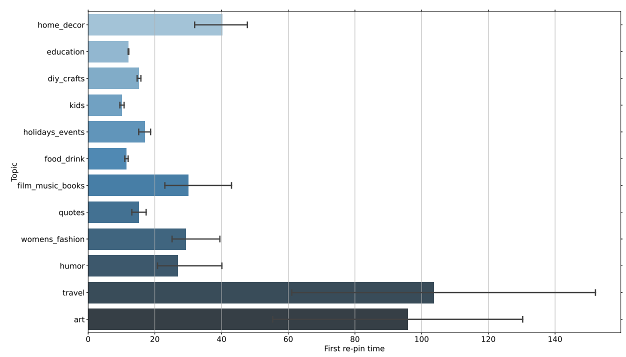


(b) Virality

Figure 6: Mean of volume and virality per topic



(a) Average re-pin time



(b) The first re-pin time

Figure 7: Median of the velocity measures for the top topics

Additionally, the pins from this domain exhibit high virality and rapid diffusion. Such active and inspiring teachers exemplify the influential role of teachers on social media and the impact they can have on their peers in the digital age. Figure 5 showcases a popular sample pin from *moffattgirls.blogspot.com*.

❏ Materials from *teacherspayteachers.com* also demonstrate high volume and virality, suggesting the popularity of educational materials from this source. Interestingly, the velocity measures for pins from *teacherspayteachers.com* reveal a short first re-pin time but a relatively longer average re-pin time. This is because pins from this popular source continue to be diffused across Pinterest over an extended period, resulting in a longer average re-pin time.

❏ Another observation is the long first re-pin time for pins from the *Uploaded by User* domain. We surmise that this may be due to the following reason: since these pins do not originate from a specific internet website (i.e., they have no domain), other users might be hesitant to save them promptly, possibly due to trust issues. However, once these pins gain (intial) popularity, they become more widespread and diffuse across the network.

5.3 Teacher Attributes and Diffusion

In addition to the resource itself, a resource producer (e.g., a teacher) can also influence the diffusion process [31]. There exists a substantial body of literature focused on identifying influential spreaders in social media, based on their attributes [32, 33, 34]. Consequently, in this section, we investigate whether teacher attributes are associated with the diffusion of the resources they curate. To achieve this, we considered the following ten teacher-related attributes and analyzed their relationship with diffusion metrics:

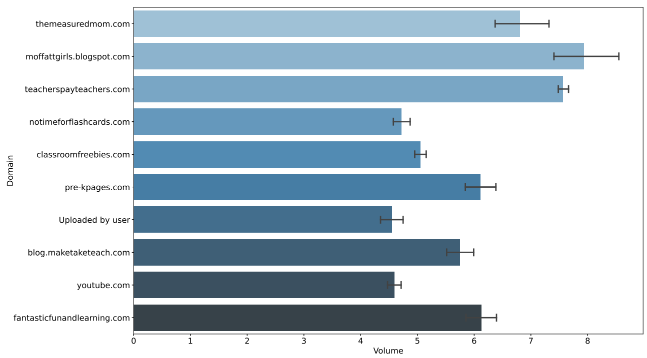
1. Number of pins: Assessing whether a teacher's activity level leads to widespread and fast diffusion of their materials.

2. Number of boards: Similar to the number of pins, this attribute also evaluates the impact of a teacher's activity level on diffusion.

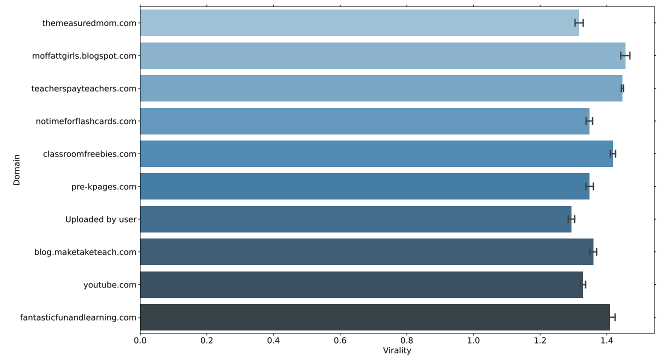
3. Number of followers: Investigating whether resources of a teacher with more followers have a higher chance of being disseminated in the network.

4. Number of followees: Examining how this attribute affects the diffusion measures.

5. Total number of friends (followers and followees combined): Analyzing the impact of this attribute on diffusion measures.

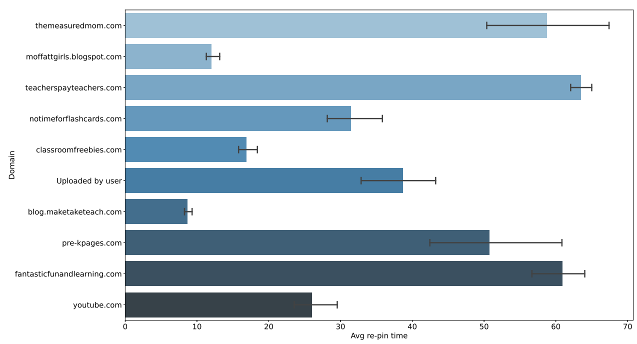


(a) Volume

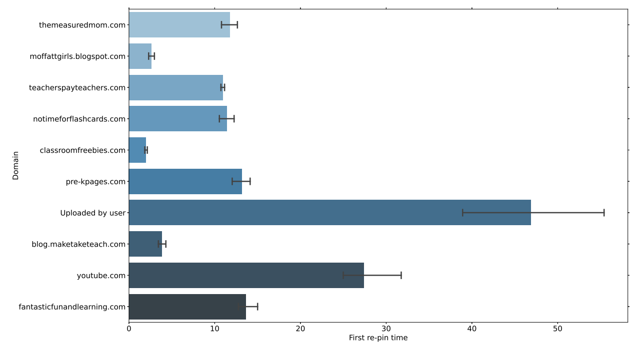


(b) Virality

Figure 8: Mean of volume and virality for the top 10 domains of teacher-curated resources



(a) Average re-pin time



(b) The first re-pin time

Figure 9: Median of the velocity measures for the top 10 domains of teacher-curated resources

6. Reciprocity: Investigating the relationship between reciprocity and diffusion, in order to determine whether having a stronger connection between a teacher and their online friends affects the diffusion.

7. Eigenvector centrality: Assessing whether resources of more central teachers have a higher chance to be adopted by other users and perhaps at a faster rate.

8. Betweenness centrality: Similar to eigenvector centrality, this attribute also evaluates the structural importance of a teacher in the network.

9. Closeness centrality: Another measure of centrality, investigating the influence of a teacher's structural importance on the diffusion of resources.

10. Local clustering coefficient: Quantifying how close a user's neighbors are to a complete graph (a clique), as previous studies [35, 36] have shown that cliques in school-level teacher networks can lead to better diffusion of information.

To explore the relationship between teacher attributes and diffusion measures, we conducted four regression analyses. In each analysis, teacher attributes served as the independent variables, while the corresponding diffusion measure acted as the dependent variable. Our goal was to determine the extent to which each teacher's attributes could explain a diffusion measure. It is important to note that we focused only on teachers who were the roots of the diffusion trees, as our aim was to identify the attributes of the pin producers, not those who further engaged in re-pinning. Consequently, a teacher could be associated with multiple diffusion trees as the root. In order to perform a teacher-level analysis, we aggregated the values of each diffusion measure for all diffusion trees associated with a teacher. For volume and virality, we calculated the mean values. For the velocity measures, we opted for the median, as it provides a more accurate estimation than the mean, as previously discussed. Finally, we utilized the statsmodels package [37] in Python to fit ordinary least squares (OLS) for each regression analysis. The results are shown in Tables 3 and 4. We make the following observations based on these results:

❏ The adjusted R-squared values are high for volume and virality, indicating that teacher attributes can significantly explain these measures. Conversely, these attributes fail to sufficiently explain the velocity measures, as reflected by the low adjusted R-squared values. This interpretation is corroborated by examining the mean squared errors: while these values are low for volume and virality, suggesting a good model fit, they are high for the velocity measures, indicating a less accurate fit.

❏ Regarding the number of pins, the coefficients for all

Table 3: Regression analysis results of predicting volume and virality using teacher attributes

Attribute	Volume Coefficient	Volume Std error	Volume t	Volume P > |t|	Virality Coefficient	Virality Std error	Virality t	Virality P > |t|
#Pins	4.449e-07	2.44e-06	0.182	0.855	-2.419e-06	2.29e-07	-10.574	0.000
#Boards	-0.0011	0.000	-3.049	0.002	-0.0001	3.31e-05	-3.861	0.000
#Followers	-0.0005	0.000	-3.590	0.000	-1.318e-05	1.41e-05	-0.936	0.349
#Followees	0.0003	0.000	2.539	0.011	-3.879e-05	9.63e-06	-4.029	0.000
#Friends	-0.0003	9.35e-05	-2.976	0.003	-5.197e-05	8.77e-06	-5.923	0.000
Reciprocity	0.2359	0.084	2.816	0.005	0.0894	0.008	11.379	0.000
Eigenvector Cent	54.8268	12.913	4.246	0.000	2.5898	1.211	2.138	0.033
Betweenness Cent	28.2263	75.308	0.375	0.708	24.0530	7.065	3.405	0.001
Closeness Cent	7.4408	0.190	39.110	0.000	3.5286	0.018	197.698	0.000
LCC	1.7113	0.283	6.039	0.000	0.6131	0.027	23.064	0.000
	Mean squared error: 2.19 Adj. R-squared: 0.539				Mean squared error: 0.04 Adj. R-squared: 0.965			

Table 4: Regression analysis results of predicting velocity using teacher attributes

Attribute	Average re-pin time Coefficient	Std error	t	P > |t|	The first re-pin time Coefficient	Std error	t	P > |t|
#Pins	-0.0015	0.000	-4.077	0.000	0.0308	0.005	-6.450	0.000
#Boards	0.1079	0.061	1.781	0.075	1.3398	0.691	1.940	0.052
#Followers	0.0096	0.022	0.440	0.660	0.3589	0.294	1.223	0.222
#Followees	0.0020	0.016	0.127	0.899	-0.3620	0.201	-1.804	0.071
#Friends	0.0116	0.014	0.822	0.411	-0.0031	0.183	-0.017	0.986
Reciprocity	-143.80	17.34	-8.290	0.000	-1005.0319	163.868	-6.133	0.000
Eigenvector Cent	-8090.16	1884.5	-4.293	0.000	-8.951e+04	2.53e+04	-3.544	0.000
Betweenness Cent	8936.45	1.1e+04	0.811	0.417	1.404e+05	1.47e+05	0.953	0.340
Closeness Cent	607.29	36.66	16.5	0.000	85375.8527	372.131	14.446	0.000
LCC	410.17	62.89	6.522	0.000	4871.5493	554.231	8.790	0.000
	Mean squared error: 85667.98 Adj. R-squared: 0.263				Mean squared error: 18323220.30 Adj. R-squared: 0.143			

measures were generally low. This suggests that a high activity rate of a pin's producer does not necessarily correlate with the diffusion of pins. This is logical, as simply saving more pins and creating more boards on Pinterest does not guarantee widespread diffusion of these resources. The only notable exception was the number of boards for the first re-pin time, where the coefficient was both positive and relatively large. This can be attributed to the fact that Pinterest users can follow a board independently, without having to follow its curator. Consequently, the more boards a user has, the higher the likelihood that someone could quickly re-pin from any of these boards. However, based on our findings, such rapid adoption does not necessarily translate to high volume and virality for the pin.

❏ The coefficients related to the number of connections, namely the number of followers, followees, and friends, were notably low. Although the coefficient of the number of followers was relatively high for the first re-pin time, it lacked statistical significance, as indicated in the P>|t| column. We posit that the low coefficient values for the number of connections can be attributed to Pinterest's nature as a social curation platform. On this website, users have the ability to re-pin resources from others without necessarily following them.

❏ Reciprocity exhibits a low coefficient for virality, yet a relatively high one for volume. Teachers with high reciprocity tend to have strong relationships with their online friends, fostering a trusting environment for re-pinning their resources. However, virality is a complex measure that cannot be sufficiently explained by reciprocity alone. As for the velocity measures, the coefficients of reciprocity are large and negative, a phenomenon that warrants further exploration.

❏ The most significant observation from this section is the relationship between the centrality metrics and the volume and virality. With the exception of betweenness centrality for volume, the centrality metrics provide a comprehensive explanation for both virality and volume. This is likely due to the centrality metrics taking into account the network's structure, a crucial factor in information diffusion. For example, a teacher with high eigenvector centrality is connected to other users with high centrality. Consequently, when these central neighbors re-pin a resource, the likelihood of wide diffusion increases due to their significant structural influence. Furthermore, both closeness and betweenness centrality are related to the shortest paths in the network, which play a critical role in the diffusion of information [38].

❏ As previously noted, the local clustering coefficient plays a pivotal role in the diffusion of information within school-level teacher networks [35, 36]. Moreover, findings from this section of the dissertation demonstrate that this attribute is equally significant in the diffusion of information across the network of teachers on Pinterest.

From the observations discussed, we can infer that teacher attributes have a substantial impact on the volume and virality of the resources they curate. Specifically, a teacher's structural characteristics at the network level play a crucial role in determining their resources' volume and virality. However, these attributes do not adequately predict the speed at which these resources are diffused.

6. CONCLUSION AND FUTURE WORK

This paper extensively analyzed the diffusion process of teacher-curated resources on Pinterest. Our first task involved constructing a comprehensive set of diffusion trees for these resources on the platform. Subsequently, we defined three critical measures to capture the essence of the diffusion process: volume, virality, and velocity. Finally, our in-depth analysis revealed that educational materials experience wide and rapid dissemination across Pinterest.

To further our understanding of diffusion dynamics, we executed multiple regression analyses to identify the teacher attributes that significantly influence the diffusion process. Our findings underscored the crucial role of structural attributes in the diffusion of teacher-curated resources on Pinterest. This is an important insight, demonstrating the relevance of network-level structural characteristics in predicting the volume and virality of resources. However, our study also indicated that these teacher attributes do not adequately explain the speed of diffusion, pointing to the complexity of the diffusion process and suggesting the need for further investigation.

Our large-scale, data-driven study not only deepens the understanding of how teacher-curated materials diffuse on Pinterest but also sets the stage for future research. The insights garnered here could be instrumental in optimizing information dissemination strategies on social curation platforms like Pinterest and beyond. By identifying the key factors that influence diffusion, educational stakeholders can harness these attributes to enhance the reach and impact of curated resources. Additionally, our findings may inspire researchers to delve deeper into the mechanisms underlying the diffusion process, encouraging the exploration of other factors and attributes we have not covered in this study. This research opens up a rich avenue for further inquiry and innovation in information diffusion in online educational networks.

There are a couple of interesting future directions:

- ➜ **In-depth Analysis of Velocity Measures:** Our study indicated that teacher attributes do not sufficiently explain the speed of diffusion (velocity). Therefore, future research could focus on a more detailed investigation into the factors influencing the velocity of resource diffusion. This might include considering additional user attributes, the nature of the content being shared, or even network-level factors such as pins' timing and user engagement dynamics over time.

- ➜ **Role of Content Characteristics:** This study primarily focused on the role of teacher attributes in the diffusion process. Future research could extend this to consider the curated resources' characteristics. This could involve analyzing the resources' content, format, topic, or even aesthetic appeal and how these factors might influence their diffusion.

- ➜ **Temporal Analysis of Diffusion:** Another interesting direction could be the temporal analysis of diffusion processes. How do teacher attributes and the diffusion of their resources evolve over time? Longitudinal studies could provide further insights into the dynamic nature of information diffusion on Pinterest.

- ➜ **Cross-platform Studies:** This research was focused on Pinterest. Future studies could examine diffusion processes on other social curation platforms or across multiple platforms. Such studies could reveal platform-specific characteristics influencing diffusion and offer a comparative perspective.

- ➜ **Impact of Algorithmic Recommendations:** Pinterest, like many other platforms, uses recommendation algorithms to suggest pins to users. Future research could explore how these algorithms influence the diffusion process. This could involve studying the interaction between user behavior and the platform's algorithmic curation in shaping diffusion patterns.

7. REFERENCES

[1] Hamid Karimi, Tyler Derr, Kaitlin T Torphy, Kenneth A Frank, and Jiliang Tang. A roadmap for incorporating online social media in educational research. *Teachers College Record*, 121(14):1–24, 2019.

[2] Hamid Karimi, Tyler Derr, Kaitlin T Torphy, Kenneth A Frank, and Jiliang Tang. Towards improving sample representativeness of teachers on online social media: A case study on pinterest. In *Artificial Intelligence in Education: 21st International Conference, AIED 2020, Ifrane, Morocco, July 6–10, 2020, Proceedings, Part II 21*, pages 130–134. Springer International Publishing, 2020.

[3] Kaitlin Torphy Knake, Hamid Karimi, Sihua Hu, Kenneth A Frank, and Jiliang Tang. Educational research in the twenty-first century: Leveraging big data to explore teachers' professional behavior and educational resources accessed within pinterest. *The Elementary School Journal*, 122(1):86–111, 2021.

[4] Hamid Karimi. *Teachers in social media: a data science perspective*. PhD thesis, Michigan State University, 2021.

[5] Hamid Karimi, Jiliang Tang, Xochitl Weiss, and Jiangtao Huang. Automatic identification of teachers in social media using positive unlabeled learning. In *2021 IEEE International Conference on Big Data (Big Data)*, pages 643–652. IEEE, 2021.

[6] Hamid Karimi, Kaitlin Torphy Knake, and Kenneth A Frank. Teachers in social media: A gender-aware behavior analysis. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 1842–1849. IEEE, 2022.

[7] Kaitlin Torphy Knake, Zixi Chen, Xiuqi Yang, and Jordan Tait. Pinterest curation and student achievement: The effects of elementary mathematics resources on students' learning over time. *the elementary school journal*, 122(1):57–85, 2021.

[8] Kaitlin Torphy, Sihua Hu, Yuqing Liu, and Zixi Chen. Teachers turning to teachers: teacherpreneurial behaviors in social media. *American Journal of Education*, 127(1):49–76, 2020.

[9] Christine Greenhow, Sarah M Galvin, Diana L Brandon, and Emilia Askari. A decade of research on k-12 teaching and teacher learning with social media: Insights on the state of the field. *Teachers College Record*, 122(6):n6, 2020.

[10] V. Darleen Opfer, Julia H. Kaufman, and Lindsey E. Thompson. *Implementation of K-12 state standards for mathematics and English Language Arts and literacy: Findings from the American Teacher Panel*. RAND Corporation, Santa Monica, CA, 2016.

[11] Statista. Number of monthly active pinterest users worldwide from 2nd quarter 2016 to 3rd quarter 2021 (in millions)., 2021.

[12] Sihua Hu, Kaitlin T Torphy, Amanda Opperman, Kimberly Jansen, and Yun-Jia Lo. What do teachers share within socialized knowledge communities: A case of pinterest. *Journal of Professional Capital and Community*, 2018.

[13] Jeffrey Carpenter, Amanda Cassaday, and Stefania Monti. Exploring how and why educators use pinterest. In *Society for Information Technology & Teacher Education International Conference*, pages 2222–2229. Association for the Advancement of Computing in Education (AACE), 2018.

[14] Eric Gilbert, Saeideh Bakhshi, Shuo Chang, and Loren Terveen. " i need to try this"? a statistical overview of pinterest. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 2427–2436, 2013.

[15] Jinyoung Han, Daejin Choi, Byung-Gon Chun, Ted Kwon, Hyun-chul Kim, and Yanghee Choi. Collecting, organizing, and sharing pins in pinterest: interest-driven or social-driven? *ACM SIGMETRICS Performance Evaluation Review*, 42(1):15–27, 2014.

[16] Daehoon Kim, Jae-Gil Lee, and Byung Suk Lee. Topical influence modeling via topic-level interests and interactions on social curation services. In *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*, pages 13–24. IEEE, 2016.

[17] Regina Kasakowskij, Thomas Kasakowskij, and Kaja Fietkiewicz. Pinterest: A unicorn among social media? an investigation of the platform's quality and specifications. In *ECSM 2020 8th European Conference on Social Media*, page 399. Academic Conferences and publishing limited, 2020.

[18] Stephanie Schroeder, Rachelle Curcio, and Lisa Lundgren. Expanding the learning network: How teachers use pinterest. *Journal of Research on Technology in Education*, 51(2):166–186, 2019.

[19] Sihua Hu, Kaitlin T Torphy, Kim Evert, and John L Lane. From cloud to classroom: Mathematics teachers' planning and enactment of resources accessed within virtual spaces. *Teachers College Record*, 122(6):n6, 2020.

[20] Yuqing Liu, Kaitlin T Torphy, Sihua Hu, Jiliang Tang, and Zixi Chen. Examining the virtual diffusion of educational resources across teachers' social networks over time. *Teachers College Record*, 122(6):n6, 2020.

[21] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

[22] Jinyoung Han, Daejin Choi, Jungseock Joo, and Chen-Nee Chuah. Predicting popular and viral image cascades in pinterest. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11, 2017.

[23] Christine Greenhow, Sarah M Galvin, and K Bret Staudt Willet. What should be the role of social media in education? *Policy Insights from the Behavioral and Brain Sciences*, 6(2):178–185, 2019.

[24] Thi Bich Ngoc Hoang and Josiane Mothe. Predicting information diffusion on twitter–analysis of predictive features. *Journal of computational science*, 28:257–264, 2018.

[25] Bo Wu and Haiying Shen. Analyzing and predicting news popularity on twitter. *International Journal of Information Management*, 35(6):702–711, 2015.

[26] Andrey A Dobrynin, Roger Entringer, and Ivan Gutman. Wiener index of trees: theory and applications. *Acta Applicandae Mathematica*, 66(3):211–249, 2001.

[27] Harry Wiener. Structural determination of paraffin boiling points. *Journal of the American chemical society*, 69(1):17–20, 1947.

[28] Kaitlin T Torphy, Diana L Brandon, Alan J Daly, Kenneth A Frank, Christine Greenhow, S Hua, and Martin Rehm. Social media, education, and digital democratization. *Teachers College Record*, 122(6):1–7, 2020.

[29] Lilian Weng, Filippo Menczer, and Yong-Yeol Ahn. Virality prediction and community structure in social networks. *Scientific reports*, 3(1):1–6, 2013.

[30] Piia Varis and Jan Blommaert. Conviviality and collectives on social media: Virality, memes, and new social structures. *Multilingual Margins: A journal of multilingualism from the periphery*, 2(1):31–31, 2015.

[31] Adrien Guille, Hakim Hacid, Cecile Favre, and Djamel A Zighed. Information diffusion in online social networks: A survey. *ACM Sigmod Record*, 42(2):17–28, 2013.

[32] Qian Li, Tao Zhou, Linyuan Lü, and Duanbing Chen. Identifying influential spreaders by weighted leaderrank. *Physica A: Statistical Mechanics and its Applications*, 404:47–55, 2014.

[33] Ling-ling Ma, Chuang Ma, Hai-Feng Zhang, and Bing-Hong Wang. Identifying influential spreaders in complex networks based on gravity formula. *Physica A: Statistical Mechanics and its Applications*, 451:205–212, 2016.

[34] Frank Bauer and Joseph T Lizier. Identifying influential spreaders and efficiently estimating infection numbers in epidemic models: A walk counting approach. *EPL (Europhysics Letters)*, 99(6):68007, 2012.

[35] Kenneth Frank, Yun-jia Lo, Kaitlin Torphy, and Jihyun Kim. Social networks and educational opportunity. In *Handbook of the Sociology of Education in the 21st Century*, pages 297–316. Springer, 2018.

[36] William Penuel, Kenneth Frank, Min Sun, Chong

Kim, and Corinne Singleton. The organization as a filter of institutional diffusion. *Teachers college record*, 115(1):1–33, 2013.

[37] Skipper Seabold and Josef Perktold. statsmodels: Econometric and statistical modeling with python. In *9th Python in Science Conference*, 2010.

[38] Anastasia Mochalova and Alexandros Nanopoulos. On the role of centrality in information diffusion in social networks. 2013.